United States Environmental Protection Agency
Office of Water (4303)
EPA 821-B-00-006
October 2000

Statistical Support Document for Proposed Effluent Limitations Guidelines and Standards for the Metal Products and Machinery Industry

Statistical Support Document for Proposed Effluent Limitations Guidelines and Standards for the Metal Products and Machinery Industry

Prepared for:

U.S. Environmental Protection Agency
Office of Science and Technology
Engineering and Analysis Division
1200 Pennsylvania Ave., NW (MC 4303)
Washington, DC 20460

Prepared by:

Science Applications International Corporation
Environmental and Health Sciences Group
Health and Environment Studies and Systems Division
11251 Roger Bacon Drive
Reston, VA 20190

October 2000

ACKNOWLEDGMENTS AND DISCLAIMER

This report has been reviewed and approved for publication by the Engineering and Analysis Division, Office of Science and Technology. This report was prepared with the support of Science Applications International Corporation under the direction and review of the EPA's Office of Science and Technology. Neither the United States Government nor any of its employees, contractors, subcontractors, or their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use of, or the results of such use of, any information, apparatus, product, or process discussed in this report, or represents that its use by such third party would not infringe on privately owned rights.

ABSTRACT

This document describes the statistical methodology used to develop effluent limitations for the Metal Products and Machinery Industry. It also presents tables of the data used to develop limits.

Table of Contents

Appendices

CHAPTER 1
OVERVIEW OF ORGANIZATION AND CONTENTS OF DOCUMENT

This document describes the statistical analyses of pollutant concentrations in effluent wastewater from metal products and machinery facilities. EPA used these statistical analyses in developing the effluent limitations guidelines and standards in the rulemaking for the Metal Products and Machinery Industry (MP&M). Details of all statistical analyses conducted and data used in the analyses to support the effluent limitations guidelines and standards for the MP&M are provided. This document is organized into six chapters, a glossary of terms (chapter 7), and five appendices. The following list summarizes the content of each chapter and appendix.

Chapter 1: Overview
— Describes the organization of the document and summarizes the contents of each chapter and appendix.

Chapter 2: Analytical Data Collection Efforts and Definition of Options
— Provides an overview of the analytical data collection efforts and defines the technology options.

Chapter 3: Description of Data Conventions
— Presents data conventions and describes data aggregation and review procedures.

Chapter 4: Statistical Methodology
— Formulates the modified delta-lognormal distribution that EPA used to derive the proposed limitations.

Chapter 5: Estimation under the Modified Delta-Lognormal Distribution
— Describes the estimation of long-term averages (LTAs) and variability factors (VFs) at the facility and pollutant levels.

Chapter 6: Derivation of the Proposed Limitations
— Describes the derivation of the proposed limitations.

Chapter 7: Glossary of Terms
— Defines technical terms used in this document.

Appendix A: Daily Effluent Data Listing
— Provides a listing of effluent daily data by subcategory and option for regulated pollutants.

Appendix B: Effluent Data Summary Statistics
— Provides summary statistics by subcategory and option for the data from each facility used to characterize the treatment in the regulated options.

Appendix C: Facility-Level Long-Term Averages and Variability Factors
— Summarizes facility-specific LTAs and VFs by subcategory and option.

Appendix D: Pollutant-Level Long-Term Averages, Variability Factors, and Limitations
— Presents pollutant-level LTAs, VF estimates, and concentration-based limitations by subcategory and option.

Appendix E: Effluent Limitations
— Summarizes the daily and monthly limitations by treatment, subcategory, and pollutant.

Appendix F: Production-based Limits for the Steel Forming and Finishing Subcategory
— Lists the daily and monthly production-based limitations by pollutant and manufacturing process for the Steel Forming and Finishing subcategory.

CHAPTER 2
ANALYTICAL DATA COLLECTION EFFORTS AND DEFINITION OF OPTIONS

2.1 EPA Wastewater Sampling

Pollutant concentration data collected during EPA wastewater sampling efforts and facility supplied data are the basis of estimates for this effluent guideline. Data from 54 sampling episodes were used to derive pollutant-specific concentration-based limitations for the following subcategories: General Metals, Metal Finishing Job Shops, Non-Chromium Anodizing, Printed Wiring Boards, Oily Wastes, Railroad Line Maintenance, Shipbuilding Dry Docks, and Steel Forming and Finishing.

EPA collected influent and effluent data from wastewater treatment systems at MP&M facilities during Phase I (1990-1993) and Phase II (1995-1999) data collection efforts. EPA used these data to develop limitations for all eight subcategories. **Table 2-1** provides a summary of the number of episodes and facilities used for limitation development by subcategory. Because EPA sampled a number of facilities two or more times, the number of facilities differs from the number of episodes shown in **Table 2-1**.

Table 2-1.
Number of Facilities and Sampling Episodes Used for Limitation Development

Subcategory	Number of Facilities			Number of Sampling Episode		
	Phase I	Phase II	Overall	Phase I	Phase II	Overall
General Metals	13	15	28	14	15	29
Metal Finishing Job Shops	2	4	6	2	8	10
Non-Chromium Anodizing	0	2	2	0	2	2
Printed Wiring Boards	0	3	3	0	3	3
Oily Wastes	1	3	4	1	4	5
Railroad Line Maintenance	0	1	1	0	1	1
Shipbuilding Dry Dock	0	3	3	0	3	3
Steel Forming and Finishing	13	15	28	14	15	29

2.2 Definition of Subcategories and Technology Options

This section defines the subcategories and technology options selected by EPA for this proposed rule. EPA has divided the MP&M point source category into eight subcategories: General Metals, Non-chromium Anodizing, Metal Finishing Job Shops, Printed Wiring Boards, Steel Forming and Finishing, Oily Wastes, Railroad Line Maintenance, and Shipbuilding Dry Dock.

Sections 2.2.1 and 2.2.2 below describe the subcategories and technology options.

2.2.1 Subcategorization Summary

General Metals
MP&M facilities that discharge metal-bearing wastewater (with or without oil-bearing wastewater) that do not belong to the Printed Wiring Board, Non-Chromium Anodizing, Metal Finishing Job Shops, or Steel Forming and Finishing subcategories. General Metals facilities usually perform manufacturing or heavy rebuilding of metal products, parts, or machines.

Non-Chromium Anodizing
Facilities that perform aluminum anodizing without the use of chromic acid or dichromate sealants in their operations.

Metal Finishing Job Shops
Job shops that perform one or more of the following: electroplating, electroless plating, anodizing, coating (chromating, phosphating, passivation, and coloring), chemical etching and milling, and printed circuit board manufacture. A job shop is defined as "a facility which owns not more than 50 percent (on an annual area basis) of the materials undergoing metal finishing" (40 CFR 433).

Printed Wiring Board
MP&M wastewater discharges from the manufacture, maintenance, and repair of printed wiring boards.

Steel Forming and Finishing
Facilities that perform MP&M operations or cold forming operations on steel wire, rod, bar, pipe, or tube. Limits for the Steel Forming and Finishing subcategory were generated using data from the General Metals subcategory because data were unavailable for the Steel Forming and Finishing subcategory.

Oily Wastes
MP&M facilities that discharge only oil-bearing wastewater and do not belong to any other MP&M subcategories.

Railroad Line Maintenance
Facilities that perform routine cleaning and light maintenance on railroad engines, cars, and car-wheel trucks and similar parts or machines. These facilities only perform MP&M unit operations defined as oily only and/or washing of final product.

Shipbuilding Dry Dock
Process wastewater generated in or on dry docks and similar structures such as building ways, graving docks, marine railways, and lift barges at shipbuilding facilities (or shipyards).

2.2.2 Technology Options Selected

In developing its technology options, EPA determined that a different set of wastewater treatment technologies was appropriate for facilities that performed unit operations that produced primarily metal-bearing wastewater than for those facilities that performed unit operations that produced primarily oily wastes. EPA concluded that the following subcategories typically produce metal-bearing wastewater (with or without associated oily-bearing wastestreams) and evaluated metals control technologies for these subcategories: General Metals, Metal Finishing Job Shops, Non-Chromium Anodizing, Printed Wiring Boards, and Steel Forming and Finishing. For the remaining subcategories (Oily Wastes, Railroad Line Maintenance, and Shipbuilding Dry Dock), EPA evaluated oily wastewater treatment technologies. EPA fully describes all of the technology options considered for each subcategory in Section 9 of the MP&M Technical Development Document.

Table 2-2 displays the selected technology options by subcategory and according to whether the option applies to existing sources or to new sources.

Table 2-2
Selected Technology Options by Subcategory

Subcategory	Selected Option for Existing Sources	Selected Option for New Sources
General Metals	Option 2	Option 4
Metal Finishing Job Shops	Option 2	Option 4
Non-Chromium Anodizing	Option 2	Option 2
Printed Wiring Boards	Option 2	Option 4
Steel Forming and Finishing	Option 2	Option 4
Oily Wastes	Option 6	Option 6
Railroad Line Maintenance	Option 10	Option 10
Shipbuilding Dry Dock	Option 10	Option 10

The following provides a brief description of the technologies included in each of the selected options.

Option 2
In-process flow control and pollution prevention, segregation of wastewater streams, preliminary treatment steps as necessary (including oils removal using oil-water separation by chemical emulsion breaking), chemical precipitation using lime or sodium hydroxide, and sedimentation using a clarifier.

Option 4
In-process flow control and pollution prevention, segregation of wastewater streams, preliminary treatment steps as necessary (including oils removal by ultrafiltration), chemical precipitation using lime or sodium hydroxide, and solids separation using a microfilter.

Option 6
In-process Flow Control, Pollution Prevention, and Oil-water separation by chemical emulsion breaking.

Option 10
In-process Flow Control, Pollution Prevention, and Oil-water separation by Dissolved Air Flotation. All of the selected options described above include the following in-process pollution prevention and water conservation technologies:

- Flow reduction using flow restrictors, conductivity meters, and/or timed rinses, for all flowing rinses, plus countercurrent cascade rinsing for all flowing rinses;
- Centrifugation and 100 percent recycling of painting water curtains; and
- Centrifugation and pasteurization to extend the life of water-soluble machining coolants reducing discharge volume by 80 percent.

CHAPTER 3
DESCRIPTION OF DATA CONVENTIONS

This chapter discusses the types of data in the MP&M analytical database and the hierarchy and procedures for aggregating multiple sampling observations within a sampling day.

3.1 Data Review

EPA's Sample Control Center (SCC) thoroughly reviewed and validated the EPA wastewater sampling data in the analytical database. During this review, the integrity of each sample was assessed to ensure that all specifications of the sampling protocol were met. In addition, engineers reviewed the data and determined that some samples should be excluded from the analyses. Samples with flags of 'T', which indicate one of the following conditions, were excluded from analyses.

- data for a treatment unit that is not a technology option
- data for a treatment system that was not operated at a proper pH for removal of targeted metals
- data for a treatment system with poor solid removal
- data for a treatment system that did not remove most of the pollutants targeted by the treatment system for removal and processed on site
- data associated with a process upset or improper sampling techniques that may impact treatment effectiveness or data quality
- data associated with specific analytes not detected in all raw wastewater influent samples to a treatment system
- data associated with specific analytes not detected in most raw wastewater influent samples to a treatment system and when detected, detected at low concentrations
- data associated with specific analytes with average raw influent concentration less than or equal to 10 times the minimum level of detection
- data associated with specific metal analytes that were not processed at the site
- data associated with specific analytes that may not have been present in a treatment system due to dilution from poor water use practices
- data associated with a treatment chemical associated with the treatment system when the chemical was not removed by the system comparably to other targeted pollutants
- data associated with specific analytes detected in the raw influent to a treatment system at low concentrations when compared to other MP&M sites or other metals processed on-site
- data for a treatment system when the concentrations of all targeted pollutants in the raw influent are low

Also during the data review, some samples were qualified with a greater than (>) sign, indicating that the reported concentration value is considered a lower limit of the actual value. This is because the reported concentration was outside the range of the analytical method. Some samples were labeled with a 'B' flag, indicating a pollutant was detected at below the minimum level of detection (sensor point). In both cases, the reported concentration values rather than the censored values were used for all computations.

3.2 Data Types

The MP&M analytical database, developed from the data reviewed and validated by SCC, contains the following three different types of samples delineated by certain qualifiers in the database:

- **Non-censored (NC):** a measured value, i.e., a sample measured above the level at which the detection decision was made.

- **Non-detect (ND):** samples for which analytical measurement did not yield a concentration above the sample-specific detection limit.

- **Right-censored (RC):** these samples were qualified with a greater than (>) sign, signifying that the reported value is considered a lower limit of the actual concentration.

3.3 Data Aggregation

Aggregation of MP&M analytical data occurred at two levels. The first level aggregated pollutant concentration data from multiple grab samples collected from a given sampling point over a specified period of time. Multiple grab samples were collected when physical compositing of the samples was impractical. The second level aggregated data from field duplicate samples. This section identifies the levels of aggregation and the methods applied to data at each level. Conventions for handling censored and non-censored data during the aggregation process are also provided.

3.3.1 Data Aggregation Across Multiple Grab Samples

Multiple grab samples are samples collected from a given sampling point over time. These samples are assigned different sample numbers. Multiple grab samples were collected when the concentration of pollutants was expected to vary during the course of sampling or when samples could not be physically composited. Within the MP&M database, Hexane Extractable Material (HEM) and Silica Gel Treated-Hexane Extractable Material (SGT-HEM) were reported as concentrations from multiple grab samples taken during one-day sampling periods. Sampling procedures require HEM and SGT-HEM to be collected as grab samples. If composited, most of the measurable amount of these analytes would remain on the composite jar due to their oily nature. Since LTAs and limitations were based on daily concentrations, multiple observations on a single day at the same sample point were averaged. Averaging analyte concentrations from multiple grab samples mathematically composites the data from the samples. The following table shows how multiple grab samples were aggregated within each sampling day/sample point combination.

Table 3-1.
Method for Averaging Multiple Grab Samples

If Observations are:	Label of "Average"	Value of "Average"
All NC	NC	$\Sigma NC_i/n$
All ND	ND	Maximum Detection Limit
NC and ND	NC	$(\Sigma NC_i + \Sigma DL_i)/n$

n = number of grab samples per day
DL = sample detection limit
NC = detected sample

3.3.2 Aggregation of Field Duplicates

Another type of data aggregation for the MP&M data was performed when flags in the database identified field duplicate samples. Field duplicates are samples collected for a particular sampling point at approximately the same time, assigned different sample numbers, and flagged as duplicates for a single episode number/sampling point. Duplicates were collected for purposes of quality assurance/quality control. **Table 3-2** presents the methods used to aggregate duplicates.

Table 3-2.
Method for Averaging Field Duplicate Samples

If Observations are:	Label of "Average"	Value of "Average"
Both NC	NC	$\Sigma NC_i/2$
Both ND	ND	Maximum Detection Limit
NC and ND	NC	(NC + ND)/2

NC = detected sample
ND = sample detection limit

If multiple grab and field duplicate samples were collected from the same sampling point, the concentration data from field duplicate samples were averaged first, then the data from the resulting duplicate sample averages were averaged as a composite sample.

Listings of summary statistics following aggregation of grabs and field duplicates are presented in **Appendix B**.

CHAPTER 4
STATISTICAL METHODOLOGY: MODIFIED DELTA-LOGNORMAL MODEL

4.1 Basic Overview of Delta-Lognormal Distribution

The lognormal distribution is often appropriate for modeling effluent data. However, the presence of NDs and very low concentration measurements in the MP&M effluent data led to the consideration of a modification to the lognormal distribution in modeling such data for several reasons. First, the lognormal model assumes that all concentration values are positively-valued. Second, the actual values of NDs are not known, though each ND has a concentration somewhere between zero and the reported detection limit. In this sense, ND measurements represent, in statistical terms, what are known as censored samples.

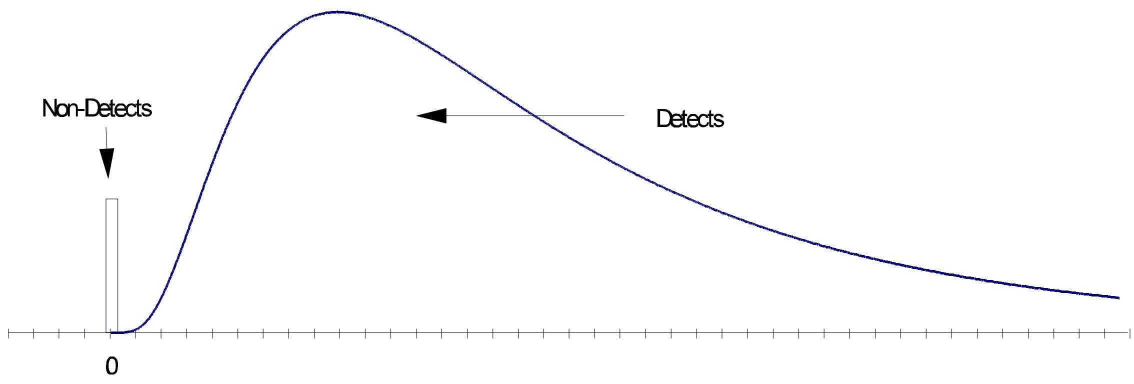
In general, censored samples are measurements for which the exact value is not known but is bounded either by an upper or lower numerical limit. Non-detects qualify in this framework as left-censored samples, which have an upper bound at the detection limit and a lower bound at zero. To model NDs as left-censored samples under a strictly lognormal density model, it is necessary to assume that the exact (but unknown) values of these measurements follow the same lognormal distributional pattern as the rest of the detected measurements and that they are positively-valued (i.e., greater than zero).

Therefore, two reasonably simple modifications to the lognormal density model have been used by EPA for several years. The first modification is known as the classical delta-lognormal model (**Figure 4-1**), first used in economic analysis to model income and revenue patterns (see Aitchison and Brown[1]). In this adaptation of the simple lognormal density, the model is expanded to include zero amounts. To do this, all positive amounts are grouped together and fit to a lognormal density. Then all zero amounts are segregated into another group of measurements representing a discrete distributional "spike" at zero. The resulting mixed distribution, combining a continuous density portion with a discrete-valued spike, is known as the delta-lognormal distribution. The delta in the name refers to the percentage of the overall distribution contained in the spike at zero, that is, the percentage of zero amounts.

---

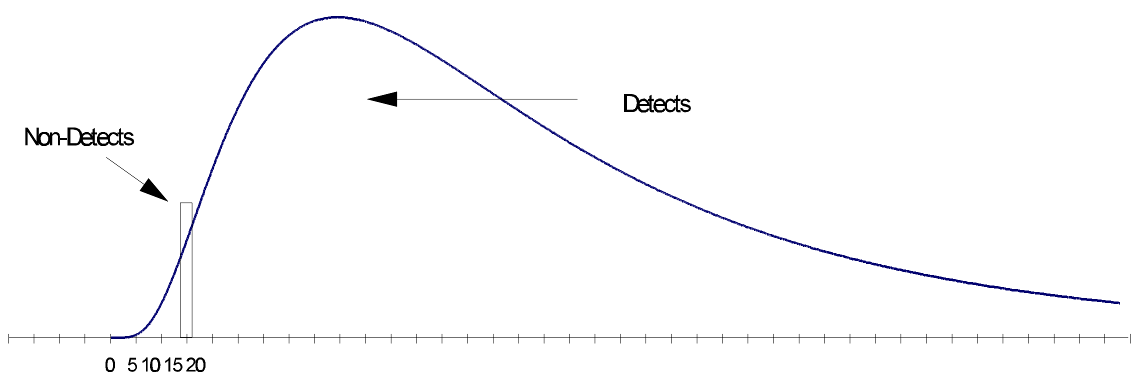
[1]. Aitchison, J. and Brown, J.A.C. (1963) The Lognormal Distribution. Cambridge University Press, pages 87-99.

Figure 4-1.
Delta-lognormal Model



Researchers at EPA (see Kahn and Rubin, 1989) further adapted the classical delta-lognormal model ("adapted model") to account for ND measurements in the same fashion that zero measurements were handled in the original delta-lognormal. Instead of zero amounts and non-zero, positive amounts, the data consisted of NDs and detects. Rather than assuming that NDs represented a spike of zero concentrations, these samples were allowed to have a single positive value, usually equal to the minimum level of the analytical method (**Figure 4-2**). Since each ND was assigned the same positive value, the distributional spike in this adapted model was located not at zero, but at the minimum level. This adaptation is appropriate since it is known that the NDs are some value greater than zero. This adapted model was used in developing limitations for the Organic Chemicals, Plastics, and Synthetic Fibers (OCPSF), Pesticides Manufacturing and Centralized Wasted Treatment rulemakings.

Figure 4-2.
Adapted Delta-lognormal Model



In the adapted delta-lognormal model, the delta again referred to those measurements contained in the discrete spike, this time representing the proportion of ND values observed within the data set. By using this approach, computation of estimates for the population mean and variance could be done easily by hand, and NDs were not assumed to follow the same distributional pattern as the detected measurements. The adapted delta-lognormal model can be expressed mathematically as follows:

$$Pr\ (U \leq u) = \begin{cases} (1-\delta)\ \Phi\ \left[(\log(u)\ -\ \mu)/\sigma\right] & if \quad 0 < u < D \\ \delta\ +\ (1-\delta)\ \Phi\ \left[(\log(D)\ -\ \mu)/\sigma\right] & if \quad u = D \\ \delta\ +\ (1-\delta)\ \Phi\ \left[(\log(u)\ -\ \mu)/\sigma\right] & if \quad u > D \end{cases} \tag{4.1}$$

where * represents the true proportion of NDs (or the probability that any randomly drawn measurement will be an ND), D equals the minimum level value of the discrete spike assigned to all NDs, M(A) represents the standard normal cumulative distribution function, and : and F are the parameters of the lognormal density portion of the model. This model assumes that all non-detected values have a single detection limit D.

It is also possible to represent the adapted delta-lognormal model in another mathematical form, one in which it is particularly easy to derive formulas for the expected value and variance of the model. In this case, a random variable distributed according to the adapted delta-lognormal distribution can be represented as the stochastic combination of three other independent random variables. The first of these variables is an indicator variable, I_u, equal to one when the measurement u is an ND and equal to zero when u is a detected value. The second variable, X_D, represents the value of an ND measurement (discrete). In the adapted delta-lognormal, this variable is always a constant equal to the concentration value assigned to each ND (i.e., equal to D in the adapted delta-lognormal model). In general, however, X_D need not be a constant, as will be seen below in the modified delta-lognormal model. The final random variable, X_C, represents the value of a detected measurement and is distributed according to a lognormal distribution (continuous) with parameters : and F.

Using this formulation, a random variable from the adapted delta-lognormal model can be written as

$$U = I_u\ X_D\ +\ (1-I_u)X_C \tag{4.2}$$

and the expected value of U is then derived by substituting the expected value of each quantity in the right-hand side of the equation. Because the variables I_u, X_D, and X_C are mutually independent, this leads to the expression

$$E(U) = \delta E(X_D)+(1-\delta)E(X_C) = \delta D\ +\ (1-\delta)\exp(\mu\ +\ 0.5\,\sigma^2) \tag{4.3}$$

where again * is the probability that any random measurement will be ND and the exponentiated expression is the familiar mean of a lognormal distribution. In a similar fashion, the variance of the adapted delta-lognormal model can be established by squaring the expression for U above, taking expectations, and subtracting the square of E(U) to get

$$Var(U) = E(U^2)\ -\ [E(U)]^2 = \delta Var(X_D)\ +\ (1-\delta)Var(X_C)\ +\ \delta(1-\delta)[E(X_D)-E(X_C)]^2. \tag{4.4}$$

Since, in the adapted delta-lognormal formulation, X_D is a constant, this expression can be reduced to the following:

$$Var(U) = (1-\delta)\exp(2\mu+\sigma^2)[\exp(\sigma^2)-(1-\delta)]\ +\ \delta(1-\delta)D[D-2\exp(\mu+0.5\sigma^2)]. \tag{4.5}$$

In order to estimate the adapted delta-lognormal mean and variance from a set of observed sample measurements, it is necessary to derive sample estimates for the parameters *, :, and F. * is typically estimated by the observed proportion of NDs in the data set. : and F are estimated using the log values of the detected samples where : is estimated using the arithmetic mean of the log-detected measurements, and F is estimated using the standard deviation of these same log values; NDs are not included in the calculations. Once the parameter estimates are obtained, they are used in the formulas above to derive the estimated adapted delta-lognormal mean and variance.

To calculate effluent limitations and/or standards, it is also necessary to estimate upper percentiles from the underlying data model. Using the delta-lognormal formulation above in equation (4.1), letting $U_{"}$ represent the 100*"th percentile of random variable U, and adopting the standard notation of z_s for the s^{th} percentile of the standard normal distribution, an arbitrary delta-lognormal percentile can be expressed as the following:

$$U_{\alpha} = \begin{cases} \exp(\mu + \sigma \; z_{\alpha/1-\delta}) & \text{if} \quad (1-\delta)\Phi((\log(D)-\mu)/\sigma) \geq \alpha \\ D & \text{if} \quad \delta + (1-\delta)\Phi((\log(D)-\mu)/\sigma) \geq \alpha \\ \exp(\mu + \sigma \; z_{\alpha-\delta/1-\delta}) & \text{if} \quad \delta + (1-\delta)\Phi((\log(D)-\mu)/\sigma) < \alpha \end{cases} \qquad \textbf{(4.6)}$$

The daily maximum limitations are established on the basis of an estimated upper 99th percentile from the underlying data model, so that 0.99 would be substituted for " in the above expression. To derive the daily VF for the 99th percentile based on the adapted delta-lognormal model, divide $U_{.99}$ in the expression above by the LTA

4.2 Motivations for Modifications to the Adapted Delta-Lognormal Model

While the adapted delta-lognormal model has been used successfully for years by EPA in a variety of settings, the model makes two key assumptions about the observed data that are not fully satisfied within the MP&M analytical database. First, the discrete spike portion of the adapted delta-lognormal model is a fixed, single-valued probability mass associated (typically) with all the ND measurements. If all ND samples in the MP&M database had roughly the same reported detection limit, this assumption would be adequately satisfied. However, the detection limits reported were sample-specific and, therefore, varied as a result of factors such as dilution. Because of this variation in detection limits, a single-valued discrete spike could not adequately represent the set of ND measurements observed in the MP&M database and a modification to the model was considered.

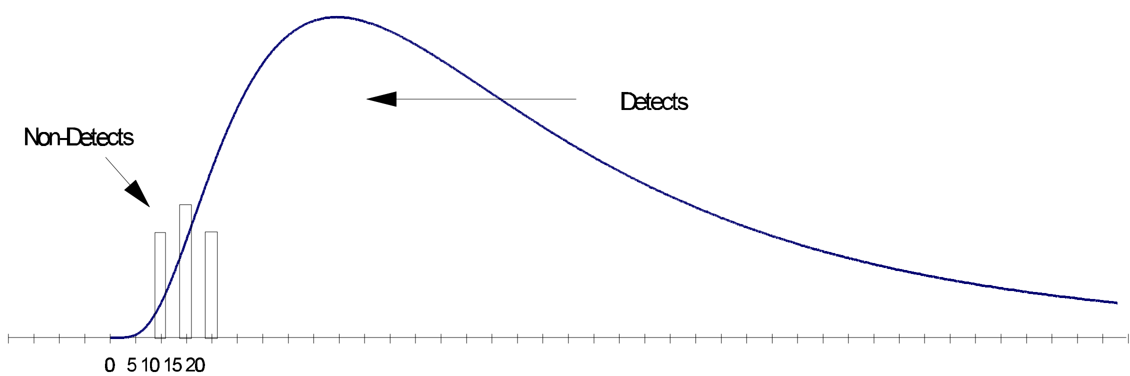
In addition, the adapted delta-lognormal model sets all NC values below the detection limit to the Minimum Level of the analytical method. For example, if the Minimum Level for N-Dodecane was 0.10 ug/l, then any NC samples reported below 0.10 ug/l were set to 0.10 ug/l. There were a few instances in the MP&M analytical studies where an NC value was reported below the Minimum Level of the analytical method.

4.2.1 Modification of the Discrete Spike

To appropriately modify the adapted delta-lognormal model for the observed MP&M database, a modification was made to the discrete, single-valued spike representing ND measurements. Because

ND samples have varying detection limits, the spike of the delta-lognormal model has been replaced by a discrete distribution made up of multiple spikes. Each spike in this modification is associated with a distinct detection limit observed in the MP&M database. Thus, instead of assigning all NDs to a single, fixed value, as in the adapted model, NDs can be associated with multiple values depending on how the detection limits vary (**Figure 4-3**).

Figure 4-3.
Modified Adapted Delta-lognormal Model



In particular, because the detection limit associated with an ND sample is considered to be an upper bound on the true value, which could range conceivably from zero up to the detection limit, the modified delta-lognormal model used here assigns each ND sample to its reported detection limit.

Once each ND has been associated with its reported detection limit, the discrete "delta" portion of the modified model is estimated in a way similar to the adapted delta-lognormal distribution, only now multiple spikes are constructed and linked to the distinct detection limits observed in the data set. In the adapted model, the parameter * is estimated by computing the proportion of NDs. In the modified model, * again represents the proportion of NDs, but is divided into the sum of smaller fractions, $*_i$, each representing the proportion of NDs associated with a particular and distinct detection limit. Thus it can be written as

$$\delta = \sum_i (\delta_i). \tag{4.7}$$

If D_i equals the value of the ith smallest distinct detection limit in the data set, and let the random variable X represent a randomly chosen ND sample, then the discrete distribution portion of the modified delta-lognormal model can be mathematically expressed as

$$Pr(X_D \leq x) = \sum_{i:D_i \leq x} \delta_i. \tag{4.8}$$

The mean and variance of this discrete distribution can be calculated using the following formulas:

$$E(X_D) = \frac{1}{\delta}\sum_i \delta_i D_i \quad \textit{and} \quad Var(X_D) = \frac{1}{\delta^2}\sum_i \sum_{i<j} \sum_j \delta_i \delta_j (D_j - D_i)^2. \qquad \textbf{(4.9)}$$

It is important to recognize that, while replacing the single discrete spike in the adapted delta-lognormal distribution with a more general discrete distribution of multiple spikes increases the complexity of the model, the discrete portion with multiple spikes plays a role in limitations and standards development identically parallel to the single spike case and offers flexibility for handling multiple observed detection limits.

CHAPTER 5
ESTIMATION UNDER THE MODIFIED DELTA-LOGNORMAL MODEL

Once the modifications to the adapted delta-lognormal distribution are made, it is possible to fit a wide variety of observed effluent data sets to the modified model. Multiple detection limits for non-detects(NDs) can be handled. The same basic framework can be used even if there are no ND values or censored data.

Combining the discrete portion of the model with the continuous portion, the cumulative probability distribution of the modified delta-lognormal model can be expressed as follows, where D_k denotes the largest distinct detection limit observed among the NDs and the first summation is taken over all those values, D_i, that are less than u.

$$Pr(U \leq u) \;=\; \begin{cases} \sum_{i:D_i<u} \delta_i + (1-\delta)\Phi\left[(\log(u)-\mu)/\sigma\right] & \text{if} \quad u<D_k \\ \delta + (1-\delta)\Phi\left[(\log(u)-\mu)/\sigma\right] & \text{if} \quad u \geq D_k \end{cases} \tag{5.1}$$

Again combining the discrete and continuous portions of the modified model, the expected value of the random variable U can be derived as a weighted sum of the expected values of the discrete and continuous lognormal portions of the distribution. This follows because the modified delta-lognormal random variable U can be expressed again as a combination of three other independent variables, that is,

$$U \;=\; I_u X_D \;+\; (1-I_u)X_C \tag{5.2}$$

where this time X_D represents a random ND from the discrete portion of the model, X_C represents a random detected measurement from the continuous lognormal portion, and I_u is an indicator variable signaling whether any particular random measurement is detected or not. Then the expected value and variance of U have forms somewhat similar to the adapted delta-lognormal model, namely

$$E(U) \;=\; \sum_i \delta_i D_i \;+\; (1-\delta)\exp(\mu + 0.5\,\sigma^2) \tag{5.3}$$

$$\begin{aligned} Var(U) \;=\; & \frac{\sum_i \sum_{j,\, i<j} \delta_i \delta_j (D_i - D_j)^2}{\delta} \;+\; (1-\delta)\exp(2\mu + \sigma^2)(\exp(\sigma^2)-1) \\ & +\; \delta(1-\delta)\left[\frac{\sum_i \delta_i D_i}{\delta} - \exp(\mu + 0.5\sigma^2)\right]^2 \end{aligned} \tag{5.4}$$

where D_i = detection limit for the i^{th} ND value
D_j = detection limit for the j^{th} ND value, where $i < j$
$*_i$ = proportion of NDs with detection limit = D_i

$*_j$	=	proportion of NDs with detection limit = D_j
*	=	proportion of all NDs
:	=	mean log concentrations of non-censored (NC) values
F	=	standard deviation of log NC values.

Example:

Consider a facility that has 10 samples with the following concentrations:

Sample number	**Measurement Type**	**Concentration (mg/L)**
1	ND	10
2	ND	15
3	ND	15
4	ND	20
5	NC	25
6	NC	25
7	NC	30
8	NC	35
9	NC	35
10	NC	40

Then the mean and variance of the log NC values are calculated as follows:

$$\mu = \frac{\sum_{i=1}^{n} \ln(x_i)}{n}$$

$$= \frac{\left(2*\ln(25) + \ln(30) + 2*\ln(35) + \ln(40)\right)}{6} = 3.44$$

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\ln(x_i) - \mu\right)^2$$

$$= \frac{1}{5} \left(2*(\ln(25)-3.44)^2) + (\ln(30)-3.44)^2 + 2*(\ln(35)-3.44)^2) + (\ln(40)-3.44)^2\right) = .0376.$$

The ND components of the variance equation are:

$D_1 = 10, *_1 = 1/10$
$D_2 = 15, *_2 = 1/5$
$D_3 = 20, *_3 = 1/10.$

As such, the variance for this example is:

$$Var(x) = \frac{\frac{1}{10}*\frac{1}{5}(10-15)^2 + \frac{1}{10}*\frac{1}{10}(10-20)^2 + \frac{1}{5}*\frac{1}{10}(15-20)^2}{\frac{2}{5}} + \left(1-\frac{2}{5}\right)\exp(2*3.44+.0376)$$

$$(\exp(.0376)-1) + \frac{2}{5}\left(1-\frac{2}{5}\right)\left[\frac{\left(\frac{1}{10}*10\right)+\left(\frac{1}{5}*15\right)+\left(\frac{1}{10}*20\right)}{\frac{2}{5}} - \exp(3.44+0.5*.0376)\right]^2 = 95.8.$$

5.1 Facility-Specific Estimates

For the purposes of estimating facility-specific LTAs and VFs, EPA chose to divide the MP&M data sets into two groups based on their size (number of samples) and the type of samples in the subset, because the computations differ for each group. The groups were defined as follows:

Group 1: Less than 2 NC (detectable) samples or less than 4 total samples. Specifically, Group 1 contained all data subsets with all NDs or only one detect. Sample-specific detection limits were substituted as the values associated with non-detectable samples.

Group 2: Two or more NC (detectable) samples and 4 or more total samples. Sample-specific detection limits were substituted as the values associated with non-detectable samples.

A listing of the daily data can be found in **Appendix A**.

5.1.1 Estimation of Facility-Specific LTAs

Data from both Group 1 and 2 were included for the computation of LTAs. The LTAs were calculated as the arithmetic average of the samples. **Appendix B** lists the summary statistics by subcategory, analyte, and option.

5.1.2 Estimation of Facility-Specific VFs

EPA developed 1-day and 4-day VFs for all regulated pollutants using Group 2 data only. Group 1 data were insufficient for estimating variability. Upper percentiles and VFs for Group 1 could not be computed using the modified delta-lognormal methodology.

For Group 2, the estimates of the parameters for the lognormal portion of the data were calculated using maximum likelihood estimation in the log-domain. Upper percentiles and VFs were calculated using these estimated parameters. Calculation of these VFs is described in Section 5.1.2.1 and 5.1.2.2.

5.1.2.1 Estimation of Facility-Specific 1-day VFs

The 1-day VFs are a function of the LTA and the 99th percentile. The 99th percentile of each data subset was calculated using the modified delta-lognormal methodology by first defining $D_0 = 0$, $*_0 = 0$, and $D_{k+1} = 4$ as boundary conditions, where D_i equals the ith smallest detection limit, and $*_i$ is the associated proportion of NDs at the ith detection limit. A cumulative distribution function, p, for each data subset was computed as a step function ranging from 0 to 1. The general form for p, for a given value c, is

$$p = \sum_{i=0}^{m} \hat{\delta}_i + (1 - \hat{\delta})\ \Phi\left[\frac{\log(c) - \hat{\mu}}{\hat{\sigma}}\right], \quad D_m \le c < D_{m+1},\ m = 0,1,\ldots k \tag{5.5}$$

where M is the standard normal cumulative distribution function. The following steps were completed to compute the estimated 99th percentile of each data subset:

1. k values of p at $c = D_m$, $m = 1, \ldots k$ were computed and labeled p_m.

2. The smallest value of m, such that p_m $ 0.99, was determined and labeled as p_j. If no such m existed, steps 3 and 4 were skipped and step 5 was computed instead.

3. $p^* = p_j - *_j$ was computed.

4. If $p^* < 0.99$, then $P_{99} = D_j$,
   else if p^* $ 0.99, then

$$\hat{P}_{99} = \exp\left[\hat{\mu} + \Phi^{-1}\left[\frac{\left(0.99 - \sum_{i=0}^{j-1} \hat{\delta}_i\right)}{(1 - \hat{\delta})}\right]\hat{\sigma}\right]. \tag{5.6}$$

5. If no such m exists, such that $p_m \geq 0.99$ (m = 1,…k), then

$$\hat{P}_{99} = \exp\left[\hat{\mu} \ + \ \Phi^{-1}\left[\frac{0.99 - \hat{\delta}}{(1 - \hat{\delta})}\right]\hat{\sigma}\right]. \tag{5.7}$$

The daily VF, VF1, was then calculated as

$$VF1 \ = \ \frac{\hat{P}_{99}}{\hat{E}(U)} \tag{5.8}$$

where

$$\hat{E}(U) \ = \ \sum_i \hat{\delta}_i D_i \ + \ (1 - \hat{\delta})\exp(\hat{\mu} + 0.5\,\hat{\sigma}^2).$$

5.1.2.2 Estimation of Facility-Specific 4-day VFs

Since Since EPA is assuming for costing purposes that some of the pollutants will be monitored weekly (approximately four times a month), EPA calculated a VF for monthly averages based on the distribution of 4-day averages. In order to calculate the 4-day VF, the assumption was made that the approximating distribution of $\bar{x}_4$, the sample mean for a random sample of four independent concentration values, also is derived from this modified delta-lognormal distribution, with the same mean as the distribution of the concentration values. The mean of this distribution of 4-day averages is

$$E(\bar{U}_4) \ = \ \delta_4 E(\bar{X}_4)_D \ + \ (1 - \delta_4)E(\bar{X}_4)_C \tag{5.9}$$

where $E(X_4)_D$ denotes the mean of the discrete portion of the distribution of the average of four independent concentration values (i.e., when all observations are not detected), and $E(X_4)_C$ denotes the mean of the continuous lognormal portion of the distribution.

First, it is assumed that the probability of detection (δ) on each of the four days is independent of that on the other days, and the non-detected values are therefore not correlated; consequently, $\delta_4 = \delta^4$. Also, because

$$E(\bar{X}_4)_D \ = \ E(X_D)$$

then

$$E(\bar{U}_4) \ = \ \delta^4 \sum_{i=1}^{k} \frac{\delta_i D_i}{\delta} \ + \ (1 - \delta^4)\exp(\mu_4 + 0.5\sigma^2_4) \tag{5.10}$$

and since $E(\bar{x}_4) = E(U)$, then

$$\mu_4 \;=\; \log\left[\frac{E(U) - \delta^3 \sum_{i=1}^{k} \delta_i D_i}{(1 - \delta^4)}\right] - 0.5\sigma^2{}_4. \tag{5.11}$$

The expression for $F^2{}_4$ was derived from the following relationship:

$$Var(\bar{U}_4) \;=\; \delta_4 Var((\bar{X}_4)_D) \;+\; (1 - \delta_4) Var((\bar{X}_4)_C) \;+\; \delta_4(1 - \delta_4)[E(\bar{X}_4)_D - E(\bar{X}_4)_C]^2. \tag{5.12}$$

Because

$$Var((\bar{X}_4)_D) = \frac{Var(X_D)}{4}, \quad E(\bar{X}_4)_D = E(X_D), \quad and \quad \delta_4 = \delta^4 \tag{5.13}$$

then

$$Var(\bar{U}_4) \;=\; \delta^4 \frac{Var(X_D)}{4} \;+\; (1 - \delta^4) Var((\bar{X}_4)_C) \;+\; \delta^4(1 - \delta^4)[E(X_D) - E(\bar{X}_4)_C]^2. \tag{5.14}$$

This further simplifies to

$$\begin{aligned} Var(\bar{U}_4) \;=\; & \frac{\delta^4 \sum_{i=1}^{k} \sum_{i<j \; j=1}^{k} \delta_i \delta_j (D_i - D_j)^2}{4\delta^2} \;+\; (1 - \delta^4)\exp(2\mu_4 + \sigma^2{}_4)\,[\exp(\sigma^2{}_4) - 1] \\ & + \delta^4(1 - \delta^4)\left[\sum_{i=1}^{k} \frac{\delta_i D_i}{\delta} \;-\; \exp(\mu_4 + 0.5\sigma^2{}_4)\right]^2 \end{aligned} \tag{5.15}$$

and furthermore,

$$\exp(\sigma^2{}_4) - 1 \;=\; \frac{\left[Var(\bar{U}_4) - \frac{\delta^2 \sum_{i=1}^{k} \sum_{i<j \; j=1}^{k} \delta_i \delta_j (D_i - D_j)^2}{4} - \delta^2(1 - \delta^4)\left[\sum_{i=1}^{k} \delta_i D_i - \delta \exp(\mu_4 + 0.5\sigma^2{}_4)\right]^2\right]}{(1 - \delta^4)\exp(2\mu_4 + \sigma^2{}_4)}. \tag{5.16}$$

Then, from (5.10) above,

$$\exp(\mu_4 + 0.5\sigma^2{}_4) = \frac{(E(\bar{U}_4) - \delta^3 \sum_{i=1}^{k} \delta_i D_i)}{(1 - \delta^4)} \;=\; \frac{(E(U) - \delta^3 \sum_{i=1}^{k} \delta_i D_i)}{(1 - \delta^4)}, \quad \because\; E(\bar{U}_4) = E(U) \tag{5.17}$$

and letting

$$\eta \;=\; E(U) - \delta^3 \sum_{i=1}^{k} \delta_i D_i, \qquad \text{then,} \quad \exp(\mu_4 + 0.5\sigma^2{}_4) \;=\; \frac{\eta}{(1-\delta^4)}. \tag{5.18}$$

Furthermore,

$$\sigma^2{}_4 = \log\left[1 \;+\; \frac{\left[Var(\bar{U}_4) - \dfrac{\delta^2 \sum_{i=1}^{k} \sum_{i<j,\, j=1}^{k} \delta_i \delta_j (D_i - D_j)^2}{4} - \delta^2 (1-\delta^4)\left(\sum_{i=1}^{k} \delta_i D_i - \dfrac{\delta\,\eta}{(1-\delta^4)}\right)^2\right]}{\dfrac{(1-\delta^4)\eta^2}{(1-\delta^4)^2}}\right]. \tag{5.19}$$

Since Var(â$_4$) = Var(U)/4, then, by rearranging terms,

$$\sigma^2{}_4 \;=\; \log\left[1 \;+\; \frac{(1-\delta^4)\,Var(U)}{4\eta^2} \;-\; \frac{(1-\delta^4)\,\delta^2 \sum_{i=1}^{k} \sum_{i<j,\, j=1}^{k} \delta_i \delta_j (D_i - D_j)^2}{4\eta^2} \;-\; \frac{\delta^2 \left[\sum_{i=1}^{k} \delta_i D_i (1-\delta^4) - \delta\,\eta\right]^2}{\eta^2}\right]. \tag{5.20}$$

Thus, estimates of $:_4$ and F_4 were derived by using estimates of $*_1, \ldots *_k$ (sample proportion of NDs at observed detection limits $D_1, \ldots D_k$), : (MLE of logged values), and F^2 (MLE logvariance multiplied by $\dfrac{n}{n-1}$ to reflect estimation from sample) in the equations above.

In finding the estimated 95th percentile of the average of four observations, four NDs, not all at the same detection limit, an average can be generated that is not necessarily equal to D_1, $D_2, \ldots$, or D_k. Consequently, more than k discrete points exist in the distribution of the 4-day averages. For example, the average of four NDs at $k = 2$ detection limits are at the following discrete points with the associated probabilities:

i	$D^*{}_i$	$\delta^*{}_i$
1	D_1	$\delta_1{}^4$
2	$(3D_1 + D_2)/4$	$4\delta_1{}^3\delta_2$
3	$(2D_1 + 2D_2)/4$	$6\delta_1{}^2\delta_2{}^2$
4	$(D_1 + 3D_2)/4$	$4\delta_1\delta_2{}^3$
5	D_2	$\delta_2{}^4$

In general, when all four observations are not detected, and when k detection limits exist, the multinomial distribution can be used to determine associated probabilities; that is,

$$Pr\left[\bar{U}_4 = \frac{\sum_{i=1}^{k} u_i D_i}{4}\right] = \frac{4!}{u_1!\,u_2!\ldots u_k!}\prod_{i=1}^{k}\delta_i^{\,u_i}. \qquad \textbf{(5.21)}$$

where u_i is the number of non-detected measurement in the data set with the D_i detection limit. The number of possible discrete points, k^*, for k = 1,2,3,4, and 5 are given below:

<u>k</u>	<u>k*</u>
1	1
2	5
3	15
4	35
5	70

To find the estimated 95th percentile of the distribution of the average of four observations, the same basic steps (described in Section 5.1.2.1) as used for the 99th percentile of the distribution of daily observations were followed, with the following changes:

1. Change P_{99} to P_{95}, and 0.99 to 0.95.
2. Change D_m to D_m^*, the weighted averages of the detection limits.
3. Change $*_i$ to $*_i^*$.
4. Change k to k^*, the number of possible discrete points based on k detection limits.
5. Change the estimates of *, :, and F to estimates of $*_4$, $:_4$, and F_4, respectively.

Then, the estimate of the 95th percentile 4-day mean VF is:

$$VF4 = \frac{\hat{P}_{95}}{\hat{E}(U)}, \qquad \because \quad E(\bar{U}_4) = E(U).$$

Appendix C displays facility-level LTAs, 1-day VFs, and 4-day VFs by subcategory, analyte, and option.

5.2 Pollutant-Specific Estimates

5.2.1 Estimation of Pollutant-Specific LTAs

After estimating the facility-specific LTAs for each pollutant and option by subcategory, as described in section 5.1.1, pollutant-specific LTAs were calculated. Pollutant-specific LTAs provide one number for all facilities within a subcategory and option. Within each subcategory and option combination, the pollutant-specific LTAs were calculated as the median of the facility-specific LTAs for that pollutant. The median is the midpoint of the values ordered (i.e., ranked) from smallest to largest. If there is an odd number of values (with n = number of values), then the value of the (n + 1)/2 ordered observation is the median. If there is an even number of values, then the two values of the n/2 and [(n/2) + 1] ordered observations are arithmetically averaged to obtain the median value.

5.2.2 Estimation of Pollutant-Specific VFs

Pollutant-specific VFs, the ratio of the 99th percentile to the mean, provide estimates of the average variability across facilities.

5.2.2.1 Estimation of Pollutant-Specific 1-day VFs

Facility-specific VFs were estimated for each pollutant by technology option and subcategory, as described in section 5.1.2.1. A pollutant-specific 1-day VF was the mean of the facility-specific daily VFs for that pollutant in the subcategory and option combination.

5.2.2.2 Estimation of Pollutant-Specific 4-day VFs

Facility-specific 4-day VFs were estimated for each pollutant by technology option and subcategory, as described in section 5.1.2.2. A pollutant-specific 4-day VF was the mean of the facility-specific 4-day VFs for that pollutant in the subcategory and option combination. The pollutant-specific 4-day VFs were used to calculate 4-day limitations, as discussed in Chapter 6.

Appendix D displays pollutant-specific LTAs, 1-day VFs, and 4-day VFs by subcategory, analyte, and option.

5.3 Transfers

A transfer occurs when all data from the original subcategory are combined with all data from the specified transfer subcategory or when there are no data from the original subcategory so it is necessary to use, or transfer, data from another subcategory. Limits are then calculated using these combined data. In most cases, EPA applies transfers when there are insufficient data in a particular subcategory for the calculation of an LTA or VF. However, in certain cases, EPA applies data transfers even when sufficient data are available to calculate the limit for the specific pollutant within the original subcategory. **Table 5-1** lists the transfers applied to the MP&M data by subcategory, analyte, and option.

Table 5-1.
Transfers

Target Subcat	Target Chemical	Option	Data Used for Limits	
			LTA	VF
ANO	MANGANESE	2	GENL	GENL
	NICKEL	2	GENL	GENL
	OIL AND GREASE (AS HEM)	2	GENL	GENL

Target Subcat	Target Chemical	Option	Data Used for Limits	
			LTA	VF
	ZINC	2	GENL	GENL
GENL	AMENABLE CYANIDE		GENL,MFJ	GENL,MFJ
	CYANIDE		GENL,MFJ,DRYD	GENL,MFJ,DRYD
	LEAD	4	PWB (option 4)	PWB (option 4)
	MOLYBDENUM	4	GENL (option 2)	GENL (option 2)
	OIL AND GREASE (AS HEM)	4	GENL (option 2)	GENL (option 2)
	TIN	2	GENL	GENL
		4	GENL	GENL,PWB
	TOTAL SULFIDE	2	OILY	OILY
		4	OILY	OILY
	TOTAL ORGANIC CARBON (TOC)	4	GENL (option 2)	GENL (option 2)
	TOP	2	DRYD,GENL,MFJ, OILY,PWB,RRL	DRYD,GENL,MFJ, OILY,PWB,RRL
		4	DRYD,GENL,MFJ, OILY,PWB,RRL (option 2)	DRYD,GENL,MFJ, OILY,PWB,RRL (option 2)
MFJ	AMENABLE CYANIDE		GENL,MFJ	GENL,MFJ
	CADMIUM	4	GENL	GENL
	CHROMIUM	4	GENL	GENL
	COPPER	4	GENL	GENL
	CYANIDE		GENL,MFJ,DRYD	GENL,MFJ,DRYD
	LEAD	4	PWB (option 4)	PWB (option 4)
	MANGANESE	4	GENL	GENL
	MOLYBDENUM	2	GENL	GENL
		4	GENL (option 2)	GENL (option 2)
	NICKEL	4	GENL	GENL

Target Subcat	Target Chemical	Option	Data Used for Limits	
			LTA	VF
	OIL AND GREASE (AS HEM)	2	GENL	GENL
		4	GENL (option 2)	GENL (option 2)
	SILVER	4	GENL	GENL
	TIN	4	GENL	GENL,PWB
	TOTAL ORGANIC CARBON (TOC)	4	MFJ (option 2)	MFJ (option 2)
	TOTAL SULFIDE	2	OILY	OILY
		4	OILY	OILY
	TOTAL SUSPENDED SOLIDS	4	GENL	GENL
	TOP	2	DRYD,GENL,MFJ, OILY,PWB,RRL	DRYD,GENL,MFJ, OILY,PWB,RRL
		4	DRYD,GENL,MFJ, OILY,PWB,RRL (option 2)	DRYD,GENL,MFJ, OILY,PWB,RRL (option 2)
	ZINC	4	GENL	GENL
OILY	TOP	6	DRYD,GENL,MFJ, OILY,PWB,RRL	DRYD,GENL,MFJ, OILY,PWB,RRL
PWB	AMENABLE CYANIDE		GENL,MFJ	GENL,MFJ
	CHROMIUM	2	GENL	GENL
		4	GENL	GENL
	COPPER	2	GENL	GENL
		4	PWB	PWB (Tin)
	CYANIDE		GENL,MFJ,DRYD	GENL,MFJ,DRYD
	LEAD	2	GENL	GENL
		4	PWB	PWB (Tin)
	OIL AND GREASE (AS HEM)	2	GENL	GENL
		4	GENL (option 2)	GENL (option 2)
	MANGANESE	4	GENL (option 4)	GENL (option 4)

Target Subcat	Target Chemical	Option	Data Used for Limits	
			LTA	VF
	NICKEL	4	GENL (option 4)	GENL (option 4)
	TOTAL ORGANIC CARBON (TOC)	4	PWB (option 2)	PWB (option 2)
	TOTAL SULFIDE	2	OILY	OILY
		4	OILY	OILY
	TOTAL SUSPENDED SOLIDS	2	GENL	GENL
		4	GENL	GENL
	TOP	2	DRYD,GENL,MFJ, OILY,PWB,RRL	DRYD,GENL,MFJ, OILY,PWB,RRL
		4	DRYD,GENL,MFJ, OILY,PWB,RRL (option 2)	DRYD,GENL,MFJ, OILY,PWB,RRL (option 2)
	ZINC	2	GENL	GENL
		4	GENL	GENL
RRL	BOD 5-DAY (CARBONACEOUS)	10	RRL	DRYD, RRL
	OIL AND GREASE (AS HEM)	10	RRL	DRYD, RRL
	TOTAL SUSPENDED SOLIDS	10	RRL	DRYD, RRL

CHAPTER 6
DERIVATION OF THE PROPOSED LIMITATIONS

This chapter describes the derivation of the proposed daily and monthly limitations. Limits were calculated as the product of a model long-term average and a model variability factor (VF). This chapter describes the methods used to derive the proposed daily and monthly concentration-based limitations.

6.1 Steps Used to Derive Concentration-Based Limitations

The derivation of the concentration-based daily and monthly maximum limitations used the pollutant-specific LTAs and respective VFs.

The following steps were used to derive the concentration-based limitations. **Appendix E** provides, by option and subcategory, listings of the concentration-based pollutant-level limitations with the LTAs and VFs used to derive the proposed limitations.

Step 1: The facility-specific LTAs and 1-day and 4-day VFs were calculated for all facilities. Calculation of VFs was performed when the facility had four or more observations with two or more distinct detected values.

Step 2: For each option in the subcategory, the median of the facility-specific LTAs and the mean of the facility-specific 1-day and 4-day VFs were calculated to provide pollutant-specific LTAs and 1-day and 4-day VFs.

Step 3: The daily limitations for a pollutant were calculated using the product of the pollutant-specific LTA and the pollutant-specific 1-day VF. Monthly average limitations were calculated using the product of the pollutant-specific LTA and the pollutant-specific 4-day VF.

6.2 Proposed Limitations

6.2.1 Daily Concentration-Based Limitations

For each technology option and subcategory, pollutant-specific daily maximum concentration-based limitations were calculated as the product of the pollutant-specific daily long-term average and the pollutant-specific daily variability factor.

6.2.2 Monthly Concentration-Based Limitations

For each technology option and subcategory, monthly or 4-day pollutant-specific daily maximum concentration-based limitations were calculated as the product of the pollutant-specific daily long-term average and the pollutant-specific 4-day variability factor.

6.3 Total Organics Parameter (TOP)

EPA defined a Total Organics Parameter (TOP) as the sum of all quantifiable concentration values greater than the nominal quantitation value of the organic pollutants. **Table 6-1** below lists the components of TOP and their nominal quantitation limits.

Table 6-1.
Calculation of Total Organics Parameter (TOP) Limit

Total Organic Parameter Pollutants that are also POCs	CAS Number	Nominal Quantitation Limit (mg/L)	Pollutant has data in the LTA database for Option 2
Acrolein	107-02-8	0.05	
Benzoic acid	65-85-0	0.05	x*
Carbon disulfide	75-15-0	0.01	x
Dibenzofuran	132-64-9	0.01	
Dibenzothiophene	132-65-0	0.01	x
Isophorone	78-59-1	0.01	
n-Hexadecane	544-76-3	0.01	x
n-Tetradecane	929-59-4	0.01	x
Aniline	62-53-3	0.01	
Chloroform (trichloromethane)	67-66-3	0.01	x
Methylene chloride (dichloromethane)	75-09-2	0.01	
Chloroethane (ethyl chloride)	75-00-3	0.05	
1,1-Dichloroethane	75-34-3	0.01	
1,1,1-Trichloroethane (methylchloroform)	71-55-6	0.01	
1,1-Dichloroethylene (vinylidene chloride)	75-35-4	0.01	x
Tetrachloroethylene (perchloroethylene)	127-18-4	0.01	
Trichloroethylene	79-01-6	0.01	
Biphenyl	92-52-4	0.01	x
p-Cymene	99-87-6	0.01	x
Ethylbenzene	100-41-4	0.01	x
Toluene	108-88-3	0.01	x
N-Nitrosodimethylamine	62-75-9	0.05	
N-Nitrosodiphenylamine	86-30-6	0.02	
Chlorobenzene	108-90-7	0.01	
2,6-Dinitrotoluene	606-20-2	0.01	

Total Organic Parameter Pollutants that are also POCs	CAS Number	Nominal Quantitation Limit (mg/L)	Pollutant has data in the LTA database for Option 2
Phenol	108-95-2	0.01	
4-Chloro-*m*-cresol (*para*chloro*meta*cresol or 4-chloro-3- methylphenol)	59-50-7	0.01	x
2,4-Dinitrophenol	51-28-5	0.05	
2,4-Dimethylphenol	105-67-9	0.01	
2-Nitrophenol (*o*-nitrophenol)	88-75-5	0.02	
4-Nitrophenol (*p*-nitrophenol)	100-02-7	0.05	
Acenaphthene	83-32-9	0.01	x
Anthracene	120-12-7	0.01	
3,6-Dimethylphenanthrene	1576-67-6	0.01	x
Fluorene	86-73-7	0.01	x
Fluoranthene	206-44-0	0.01	
2-Isopropylnaphthalene	2027-17-0	0.01	x
1-Methylfluorene	1730-37-6	0.01	x
2-Methylnaphthalene	91-57-6	0.01	x
1-Methylphenanthrene	832-69-9	0.01	x
Naphthalene	91-20-3	0.01	x
Phenanthrene	85-01-8	0.01	x
Pyrene	129-00-0	0.01	x
Benzyl butyl phthalate	85-68-7	0.01	
Dimethyl phthalate	131-11-3	0.01	
Di-n-butyl phthalate	84-74-2	0.01	
Di-n-octyl phthalate	117-84-0	0.01	
Di(2-ethylhexyl) phthalate	117-81-7	0.01	x
Sum of nominal quantitation limits for pollutants that are not in the LTA database			0.47

* x indicates that the pollutant has data in the LTA database for Option 2.

EPA used the following steps to calculate the limit for TOP:

- Determine the LTA for each organic component.
- Sum the component LTAs.
- Multiply the total LTA by the mean VF across the individual organics.
- Add in the sum of nominal quantitation limits for pollutants that are not in the LTA database.

6.4 Production-based Limits for Steel Forming and Finishing

EPA calculated production-based limits for the Steel Forming and Finishing subcategory as follows:

$$Limit_{pn} = \frac{X mg}{L} \times \frac{Y gal}{short\ ton} \times \frac{8.3454\ L \times lb}{10^6\ gal \times mg} \times \frac{short\ ton}{2 \times 1000\ lb} = 0.00000417\ XY \frac{lb}{1000\ lb} = 0.00000417\ XY \frac{kg}{kkg}$$

where X is the concentration-based limit in mg/L and Y is the production value in gallons per ton. The production-based limits for Steel Forming and Finishing are listed in **Appendix F**.

CHAPTER 7
GLOSSARY OF TERMS

Censored value: a measurement value known to be in a certain range but for which the exact value is unknown (or deliberately ignored).

Daily maximum limit: daily effluent discharge limit on the amount of pollutant that may be discharged in a single day. In calculating this limit, EPA uses statistical methodologies that account for reasonable excursions from the long-tem average in a well-designed and operated treatment system. Numerically, for MP&M, the daily maximum is usually estimated as the product of the pollutant-specific LTA and the pollutant-specific daily Variability Factor (one-day VF).

Detection limit: the sample-specific value representing the lowest concentration that can be reliably distinguished from zero.

Expected value: the expected value of a function of variate values is its mean value in repeated sampling.[2]

Facility-specific Long-Term Average: average treated pollutant levels achieved over a period of time by the facility. For MP&M, the facility-specific LTA is usually computed as the arithmetic mean of all individual pollutant measurements in a given facility.

Facility-specific Variability Factor: a ratio that expresses the relationship between the average treatment performance level from the facility and an upper bound on large values that would be expected to occur only on rare occasions in a well-designed and operated treatment system.

Field Duplicates: one or more samples collected for a particular sampling point at the same time, or approximately the same time, assigned different sample numbers, and flagged as duplicate for a single episode number.

Four-day Variability Factor: a four-day average of the facility-specific or pollutant-specific variability factors.

Grab samples: one or more samples collected for a particular sampling point over time, assigned different sample numbers, and not physically composited.

Limitation: any restriction, including schedules of compliance, established by a State or the Administrator on quantities, rates, and concentrations of chemical, physical, biological, and other constituents which are discharged from point sources into navigable waters, the waters of the contiguous zone, or the ocean. (CWA Sections 301(b) and 304(b).)

---

[2] Kendall, M.G., and W.R. Buckland, 1982. A Dictionary of Statistical Terms. 4th Edition. Longman Group Ltd. New York; 213 p.

Lognormal distribution: a distribution whose set of values in logarithmic scale follow a mathematical function known as the normal distribution. The lognormal distribution is often appropriate for modeling environmental data.

LTA: Long-term average. For purposes of the effluent guidelines, average pollutant levels achieved over a period of time by a facility, subcategory, or technology option. LTAs are used in developing the limitations and standards in a proposed or final regulation.

Minimum level: the lowest concentration that can be reliably measured by the analytical method.

MLE: Maximum Likelihood Estimate. The value of an estimate of a population parameter that maximizes the likelihood of the sample.

Monthly average limit: monthly effluent discharge limit on the amount of pollutant that may be discharged on average during a month period. In calculating this limit, EPA uses statistical methodologies that account for reasonable excursions from the long-tem average in a well-designed and operated treatment system. Numerically, for MP&M, the monthly average limit is usually estimated as the product of the pollutant-specific LTA and the pollutant-specific monthly Variability Factor (four-day VF).

Mutually independent: two events are independent if the probability of one is not affected by the occurrence of the other.

Non-censored (NC): a measurement result reported as a numerical value.

Non-detect (ND): samples below the level that can be reliably measured by the analytical method. This is also known, in statistical terms, as left-censored; i.e., value having an upper bound at the sample-specific detection limit and a lower bound at zero.

One-day Variability Factor: daily average of the facility-specific or pollutant-specific variability factors.

Parameter: Numerical descriptive values that characterize populations of measurements; e.g., a population mean value.

Pollutant-specific Long-Term Average: average pollutant levels achieved by the facility. For MP&M, the pollutant-specific LTA is computed as the median of the facility-specific LTAs.

Pollutant-specific Variability Factor: expresses the relationship between the average pollutant level and an upper bound on large pollutant values that would be expected to occur only on rare occasions in a well-designed and operated treatment system. In mathematical terms, for MP&M this is the median of the facility-specific variability factors for that pollutant.

Right-censored (RC): samples qualified with a greater than (>) sign, signifying that the reported value is considered a lower limit of the actual concentration.

<u>Single-valued probability mass</u>: a statistical term sometimes used to describe the magnitude of a probability or the relative frequency of observations located at a particular variate value, as distinct from being distributed over a mathematically continuous range.

<u>Variability Factor</u>: used in calculating a limitation (or standard) to allow for reasonable, normal variation in pollutant concentrations when processed through well designed and operated treatment systems. Variability factors account for normal fluctuations in treatment. By accounting for these reasonable excursions about the long-term average, EPA's use of variability factors results in limitations that are generally well above the actual long-term averages.

<u>Weighted sum</u>: a sum of quantities to which have been attached a series of weights in order to make proper allowance for their relative importance.[1]

---

[1] Kendall, M.G., and W.R. Buckland. 1982. *A Dictionary of Statistical terms*. *4th Edition*. Longman Group Ltd. New York; 213 p.

LIST OF ABBREVIATIONS AND ACRONYMS

ANO: Non-Chromium Anodizers

CN: Cyanide

DRYD: Shipbuilding Dry Dock

GENL: General Metals

LTA: Long-term average

MFJ: Metal Finishing Job Shops

NC: Non-censored

ND: Non-detect

OILY: Oily Only

PWB: Printed Wiring Board

RC: Right-censored

RRL: Railroad Line Maintenance

SFF: Steel Forming and Finishing

VF: Variability Factor

Table of Direct and Indirect Episodes

Episode	Discharge Status
1197A	Indirect
4011	Indirect
4079	Indirect
4274	Indirect
4277	Indirect
4278	Indirect
4279	Indirect
4310	Indirect
4330	Indirect
4384	Indirect
4415	Indirect
4417	Indirect
4438	Indirect
4460	Indirect
4460A	Indirect
4470	Indirect
4471	Indirect
4737	Direct
4761	Indirect
4762	Indirect
4788	Indirect
4805	Direct
4806	Indirect
4807	Direct
4811	Indirect
4817	Indirect
4828	Indirect
4833	Direct
4834	Indirect
4847	Indirect
4851	Indirect
4854	Indirect
4855	Indirect
4856	Indirect
4866	Indirect
4867	Indirect
4869	Indirect
4871	Indirect
4872	Indirect
4876	Indirect
4877	Indirect
4882	Indirect
4883	Indirect
4891	Indirect
4892	Indirect
4893	Indirect
4894	Indirect
4904	Indirect
6048	Indirect
6178	Indirect
6179	Indirect
6186	Indirect
6187	Indirect

Episodes 4833 and 4807 are Direct for Option 2 only